

Análise de Dados com Base em
Processamento Massivo em
Paralelo

Mercado de Engenharia de Dados

André Perez

Especialista em Engenharia de Dados @ SumUp



CeMEAI
CEPID - Centro de Ciências
Matemáticas Aplicadas à Indústria

Agenda

- Breve linha do tempo
- Ecossistema
- Engenheira(o) de Dados
- Link

Mercado

Especialista em Engenharia de Dados

SumUp | Stone Co. | Serasa Experian | Amdocs

Academia

Engenheiro Eletricista @ EESC/USP

Mestrando em Ciência de Dados (interrompido) @ ICMC/USP

Extra

Cursos, lives, podcasts, artigos e projetos open source



Breve linha do tempo

1970 ~ 2000

- Poucos *players*;
- Software de código proprietário;
- Escala vertical de soluções;
- Infraestrutura *on-premise*.

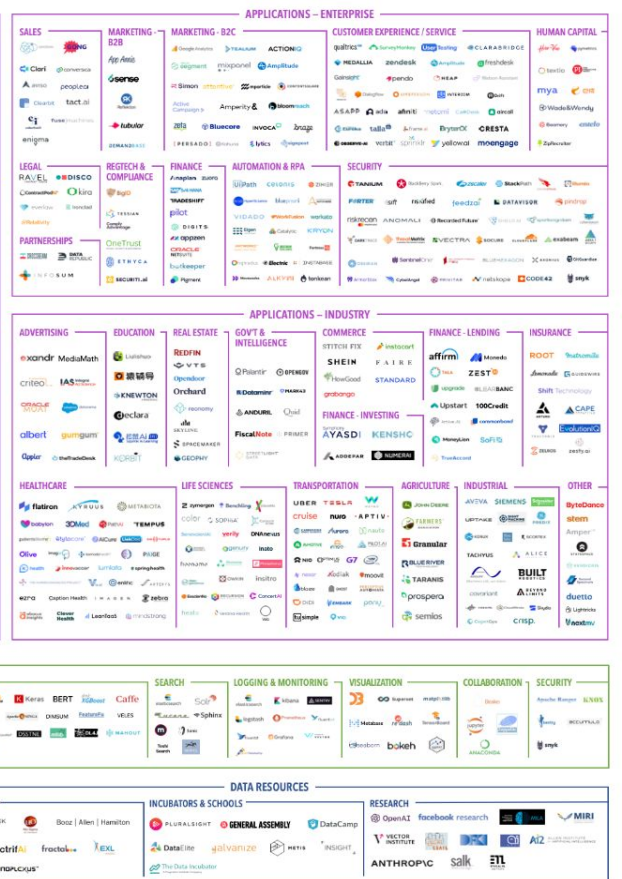
2010 ~ hoje

- Muitos *players*;
- Software de código aberto;
- Escala horizontal das soluções;
- Infraestrutura *on-premise* e *cloud*.



2000 ~ 2010

- Poucos *players*;
- Software de código aberto;
- Escala horizontal das soluções;
- Infraestrutura *on-premise*.



OPEN SOURCE

- Serviços gerenciados;
- Consultoria;
- Suporte 24/7;
- Treinamento;
- Certificação;
- etc.



Ecosistema de Dados em 2021 ([fonte](#))

Nuvem pública:

- AWS, GCP e Azure.

Processamento distribuído:

- Databricks (Spark) e DBT.

Armazenamento:

- Snowflake e MongoDB.

ETL + ELT:

- Confluent, Fivetran e Astronomer.

Plataforma de ML:

- Datarobot e Daitaku.

Operação de dados:

- Datadog.



Automação



Matt Turck
@mattturck

...

2013: Big Data

2017: ML/AI

2020: Automation

Not overhyped fads. Different stages of the same, massive secular trend.

5:58 PM · 6/14/20 · [Twitter for iPhone](#)

||| [View Tweet activity](#)

Especialização do trabalho

Engenheira(o)

DE

Cientista

DS

Analista

DA

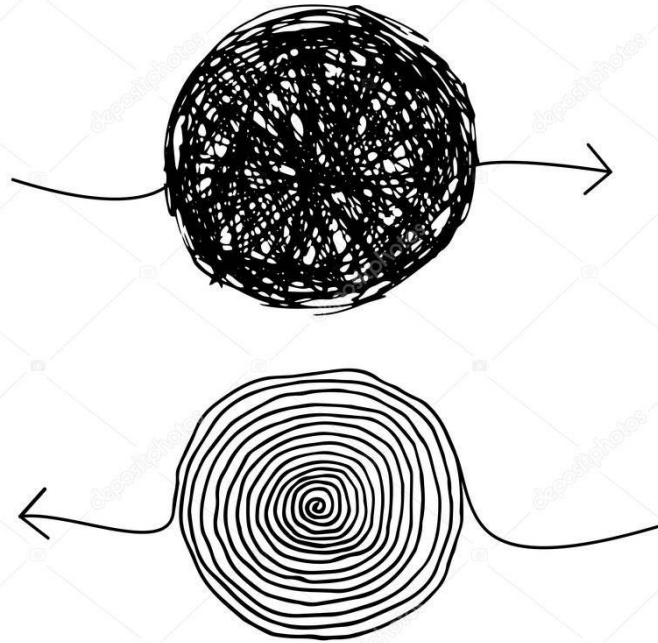
Negócio



Tecnologia



Engenheira(o) de Dados



Engenheira(o) de Dados x Cientista de Dados

2 a 5

Estima-se que são necessários de 2 a 5 engenheiras(os) de dados para cada cientista de dados.

Fonte: <https://www.oreilly.com/radar/data-engineers-vs-data-scientists/>

Engenheira(o) de Dados x Cientista e Analistas

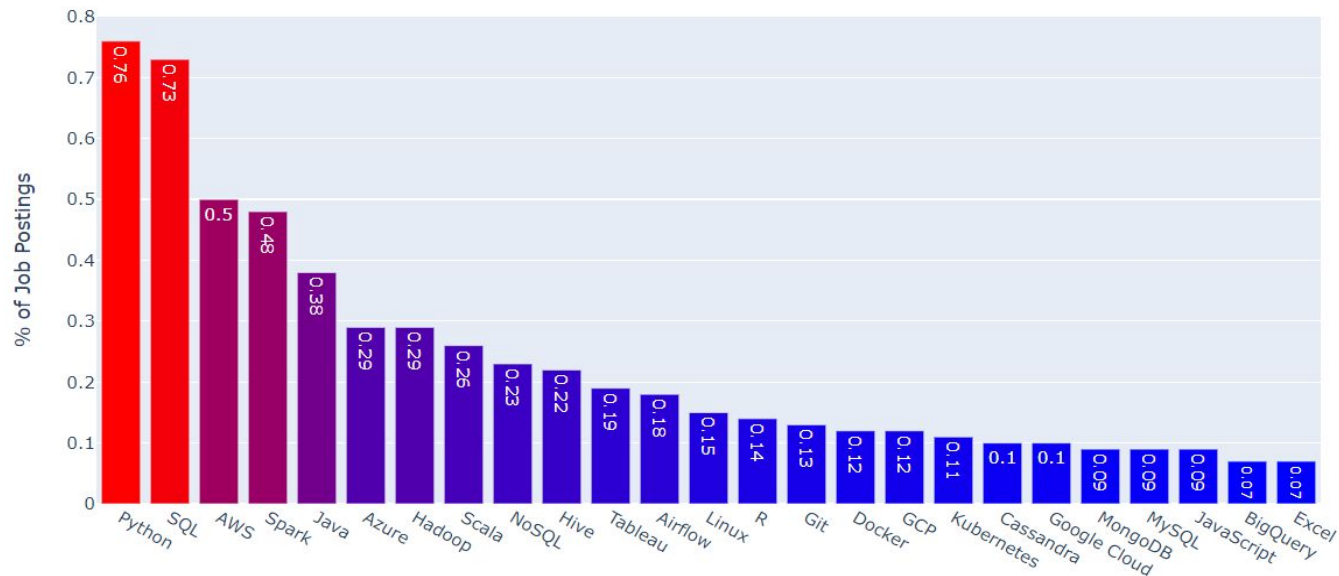
1 para 10

Na Stone Co. (antigo emprego) a razão engenheira(o) de dados para outras profissões (cientistas e analistas) era de 1 para 10.

Fonte: Fonte própria do autor.

Engenheira(o) de Dados x Cientista de Dados

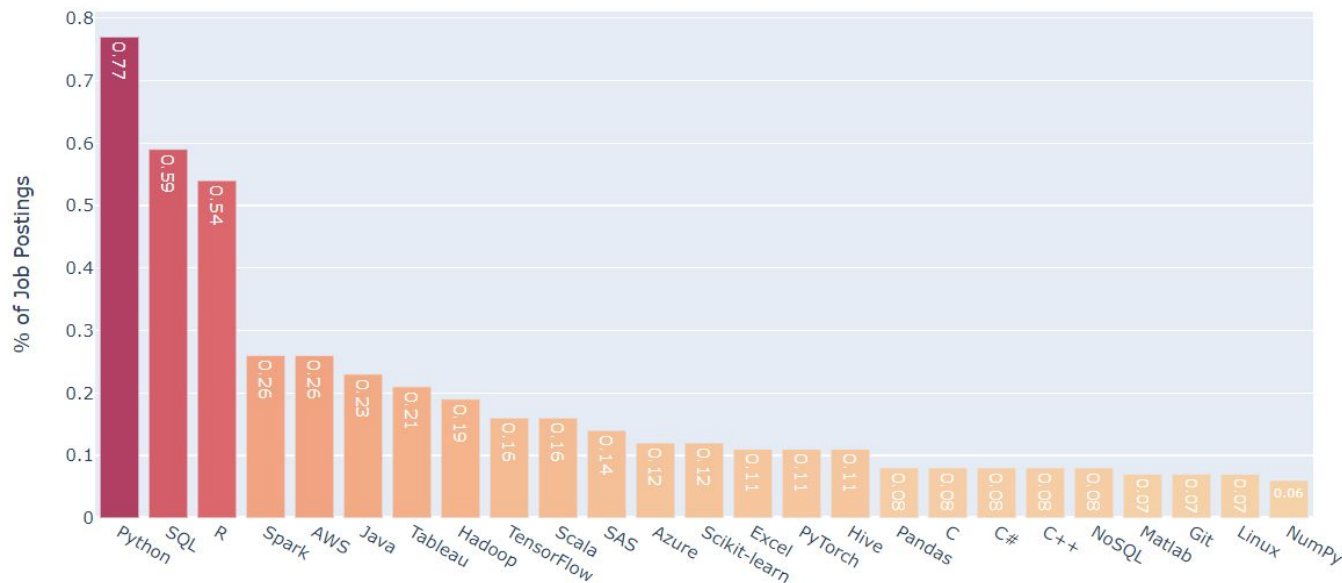
25 Most In-Demand Data Engineering Skills in 2021



Fonte: <https://www.kdnuggets.com/2021/05/most-demand-skills-data-engineers-2021.html>

Engenheira(o) de Dados x Cientista de Dados

25 Most In-Demand Data Science Skills in 2021



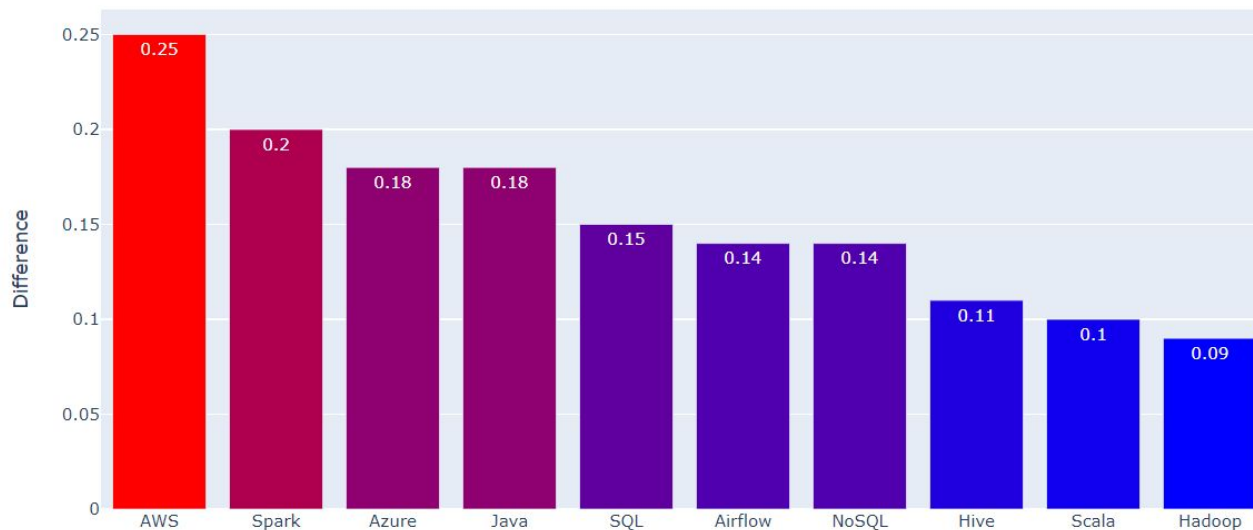
Fonte: <https://www.kdnuggets.com/2021/04/most-demand-skills-data-scientists.html>

SQL



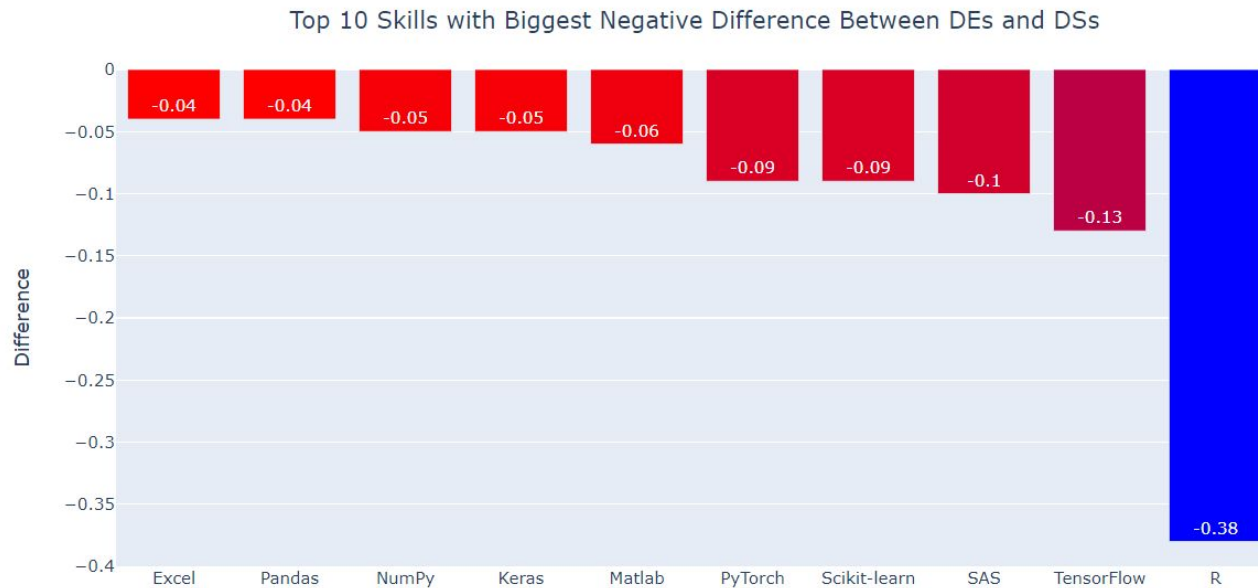
Engenheira(o) de Dados x Cientista de Dados

Top 10 Skills with Biggest Positive Difference Between DEs and DSs



Fonte: <https://www.kdnuggets.com/2021/05/most-demand-skills-data-engineers-2021.html>

Engenheira(o) de Dados x Cientista de Dados



Fonte: <https://www.kdnuggets.com/2021/05/most-demand-skills-data-engineers-2021.html>

Links

Artigos

- [The Rise of the Data Engineer](#) (Medium - 20/01/17)
- [A beginner's Guide to Data Engineering](#) (Medium - 08/01/18)

Podcasts

- [O que é Data Lake](#) (Data Hackers - 04/09/20)
- [Engenharia de Dados no Grupo Zap](#) (Data Hackers - 20/09/19)
- [Tecnologias no Nubank: 3 anos depois](#) (Hipsters.Tech - 28/05/19)
- [Nubank Data Engineering with Sujith Nair](#) (SED - 06/02/20)
- [Inteligência artificial e o nosso dia-a-dia](#) (DevPira - 05/05/21)

Live

- [Conheça o time de dados da Stone Co.](#) (EstatIDados - 08/05/21)